云 南 植 物 研 究 2009，31 (3)：227～238
Acta Botanica Yunnanica DOI：10.3724 SP.J.1143.2009.09006

拟南芥 PHD-finger 蛋白家族的全基因组分析

张 帆，杨仲南，张 森
(上海师范大学生命与环境科学学院，上海 200234)

摘要：PHD-finger 蛋白是一类广泛存在于真核生物中，在基因转录和染色质状态调控方面有重要作用的锌指蛋白。目前在动物中对 PHD-finger 蛋白的结构和功能方面的研究较为广泛和深入，而在植物中仅有少数 PHD-finger 蛋白的功能被阐明。通过 SMART 和 Pfam 等数据库分析，我们发现拟南芥中共有 70 个 PHD-finger 蛋白，其中大部分 PHD-finger 蛋白的功能未知。本文通过生物信息学分析获得拟南芥 PHD-finger 家族较为全面的信息，包括基因结构、染色体定位、基因表达、蛋白结构域、系统进化关系等，为深入研究 PHD-finger 家族蛋白的结构与功能提供了参考。
关键词：PHD-finger；拟南芥；基因家族
中图分类号：Q 943.2 **文献标识码**：A **文章编号**：0253-2700 (2009) 03-227-12

Genome-wide Analysis of PHD-finger Protein Family in *Arabidopsis thaliana*

ZHANG Fan，YANG Zhong-Nan，ZHANG Sen**
(*College of Life and Environment Sciences*，*Shanghai Normal University*，Shanghai 200234，China)

Abstract：PHD-finger proteins were found universally in eukaryotes and known as key players in regulating chromatin or transcription. In animals，many PHD-finger proteins had been well studied on structure and function. Compared to animals，only a few of plant PHD-finger proteins had been characterized functionally. By both SMART and Pfam searches，70 members of the PHD-finger family were identified in *Arabidopsis thaliana*. Most of the PHD-finger proteins in *Arabidopsis thaliana* were functionally unknown. A complete overview of this family by bioinformatic approach were presented in this paper，including the gene structure，gene localization in genome，gene expression，protein motif organization，and phylogenetic relationship of each family member. These results should provide useful information for the further structural and functional studies of PHD-finger proteins.
Key words：PHD-finger；*Arabidopsis thaliana*；Gene family

锌指蛋白是一类具有“指状”结构域的蛋白，在真核生物中数量较多，分布广泛，蛋白中的“指状”结构域富含半胱氨酸或组氨酸，保守的半胱氨酸残基和组氨酸残基通过结合锌离子以稳定整个结构域正常的空间结构，根据保守半胱氨酸残基及组氨酸残基的排列特征人们已经划出 RING，LIM，PHD 等类型（Talatsuji，1998）。PHD-finger 蛋白是真核生物中较为常见的一类具有一个或几个 PHD-finger 结构域的锌指蛋白，PHD-finger 结构域由约 60 个氨基酸组成，具有 Cys4-His-Cys3 锌结合基序特征（Kaadige and Ayer，2006），半胱氨酸残基之间以及半胱氨酸与组氨酸之间的碱基数是相对保守的（Bienz，2006），且最后一对半胱氨酸前的第 2 个氨基酸残基通常

基金项目：国家自然科学基金项目 (30671127)，上海师范大学科研项目 (sk200827)
通讯作者：Author for correspondence；E-mail：senzhang@shnu.edu.cn
收稿日期：2009-01-09，2009-03-20 接受发表
作者简介：张帆 (1982-) 男，在读硕士研究生，主要从事植物基因功能研究工作。

为色氨酸等芳香族氨基酸（Aasland 等，1995）。PHD-finger 的三维结构大致相同，为球状结构域 (globular domain)（Nakamura 等，2007），第 1、2、6 个半胱氨酸残基及组氨酸残基与一个锌离子结合，而第 3、4、7、8 个半胱氨酸残基与另一个锌离子结合，并在第 2 个和第 3 个半胱氨酸之间以及第 6 个和第 7 个半胱氨酸之间各形成一个环 (Loop) 结构（Townsley 等，2004）。

自从 Schindler 等最早在拟南芥的 HAT3.1 蛋白和玉米的 Zmhox1a 蛋白中发现 PHD-finger 结构域之后，越来越多的 PHD-finger 蛋白被发现存在于包括酵母、果蝇、人等在内的众多真核生物中，如：酵母的 YNG1、NTO1、PHO23（Martin 等，2006）等蛋白；果蝇的 ASH1（Papoulas 等，1998)、dMi-2（Kehle 等，1998）等蛋白；人的 ATRX（Gibbons 等，1997）、CBP（Ogryzko 等，1996）等蛋白。在对该家族蛋白的功能研究中，人们发现很多 PHD-finger 蛋白借助 PHD-finger 结构域和核小体的组蛋白结合，从而调节染色质状态，如：酵母 YNG 家族的蛋白 Yng1、Yng2 和 PHO23 的 PHD 结构域都具有很强的 H3K4me2 3 结合活性（Shi 等，2007），人的 BPTF（Wysocka 等，2006）等同样也和 H3K4me2 3 结合，而 BHC80 蛋白的 PHD-finger 结构域则特异识别结合无甲基化的 H3K4 (即 H3K4me0) (Lan 等，2007)，人的 ISWI SNF-染色质重构复合物（ISWI SNF-chromatin remodeling complex）的一个亚单位 ACF1 借助 PHD-finger 结构域结合到 4 个核心组蛋白的中心区域，并增加 ISWI 引起的核小体滑行的效率（Eberharter 等，2004）等。也有的 PHD-finger 蛋白借助 PHD-finger 结构域结合组蛋白以外的蛋白亚基，如果蝇的转录激活蛋白 Pygopus 的 PHD-finger 结构域结合 Legless BCL9（Kramps 等，2002），人的转录抑制蛋白 KAP-1 的 PHD-finger 结构域结合 Mi-2α 亚基的一个异构体（Schultz 等，2001）等。

在拟南芥的 PHD-finger 蛋白的研究中，现已发现甲基化胞嘧啶结合蛋白 VIM1 等能与组蛋白结合，参与染色质状态的调节（Woo 等，2007）；Trithorax 同源蛋白 ATX1 和 ATX2 等具有组蛋白的甲基转移酶活性，能调节某些基因的转录（Saleh 等，2008）；花粉和绒毡层发育的相关蛋白 MS1，可能调控小孢子生物合成中的某些基因的转录（Yang 等，2007）；小孢子母细胞减数分裂相关基因 MMD1，可能调控减数分裂过程中相关基因的转录（Yang 等，2003）等。然而，目前的研究仅阐明了少数拟南芥 PHD-finger 蛋白的功能，大多数拟南芥 PHD-finger 蛋白的功能仍不知晓。为了更加深入的了解拟南芥 PHD-finger 家族蛋白的结构和功能，对这一家族成员进行确定和全面分析将很有意义。冯英等（2004）曾从拟南芥基因组中鉴定出 45 个 PHD-finger 蛋白基因，根据蛋白包含的 PHD-finger 结构域数将其分成 2 组，并对拟南芥和水稻中的 PHD-finger 蛋白进行了序列比对及系统进化分析。但我们的研究发现从 SMART 和 Pfam 数据库中得到的拟南芥 PHD-finger 蛋白不完全一致，并且两个数据库中的信息与冯英等人的结果也存在较大差异。为更加全面的了解拟南芥中 PHD-finger 蛋白的信息，我们提取了 SMART 和 Pfam 数据库中一致的蛋白序列，并对这些序列进行筛选，最终鉴定到 70 个非冗余的拟南芥 PHD-finger 家族成员，并对这些成员做了蛋白序列比对、进化树构建，分析了这些成员的基因结构、表达、染色体定位等，为深入了解 PHD-finger 蛋白家族的进化关系及功能研究提供参考。

1 材料与方法

1.1 拟南芥中 PHD-finger 蛋白序列数据获得

从 SMART 数据库（http: smart.embl-heidelberg.de）中搜索获得拟南芥中非冗余的所有可能的 PHD-finger 结构域序列，利用 Pfam 结构域分析工具（http: pfam.sanger.ac.uk search ? tab = searchDomainBlock）对得到的这些 PHD 结构域序列进行进一步鉴定，得到候选的 PHD-finger 结构域序列，再根据 Aasland 等人推测的 PHD-finger 共有序列特征 (Bienz，2006）对候选序列进行了进一步的筛选，获得最终的 PHD-finger 结构域序列。通过 BLASTp 的方法我们在 TAIR（http: www.arabidopsis.org）中找到 PHD-finger 结构域对应的蛋白编码基因的完整信息，包括编码基因的位置、结构、大小及 PHD-finger 结构域处内含子分布等。

1.2 结构域序列分析及进化树构建

利用 ClustalX1.83 分别对预测的 PHD-finger 结构域的氨基酸序列和 PHD-finger 蛋白的氨基酸序列做多序列比对，对比中使用默认参数，并对比对结果做了手工调整。利用 BOXSHADE（http: www.ch.embnet.org software BOX_ form.Html）突出保守或相似的氨基酸残基的位置。

根据结构域序列比对结果，利用 ClustalX1.83 分别构建拟南芥 PHD-finger 结构域序列和 PHD-finger 蛋白全序列的无根系统进化树。系统进化树的构建采用邻接法（neighbor-joining），自举值（bootstrap）取 1000 次重复。

1.3 基因在染色体上的位置及倍增关系分析

利用 TAIR 网站的染色体图工具（Chromosome Map Tool）作染色体图，将预测的所有 PHD-finger 蛋白基因定位到 5 条染色体上的相应位置，对于染色体图工具不能识别的 *At1g32810* 和 *At2g25170* 两个基因，按照 TIGR（http：www.tigr.org tdb e2k1 ath1 LocusNameSearch.shtml）提供的基因序列位置信息手工将其定位到染色体上相应位置。同时利用“Paralogous in *Arabidopsis thaliana*”（http：wolfe.gen.tcd.ie athal dup）分析拟南芥 PHD-finger 家族成员中位于基因组倍增片段上的基因。

1.4 基因结构分析

首先利用 MEME（Multiple Em for Motif Elicitation）对拟南芥 PHD-finger 家族的 70 个成员做了全面分析，了解不同成员之间的共有基序。再利用 SMART 以及 Pfam 等分析工具对这 70 个蛋白中已知的结构域进行分析和注释。

1.5 基因的表达分析

利用 MPSS 数据库（http：mpss.udel.edu at）和 NCBI 的 EST 数据库分析 PHD-finger 蛋白基因的表达，主要分析在花序、果荚、叶子和根中的表达情况。对于两个数据库中不一致的信息，我们假设两个数据库中的表达信息都有效，将二者结果进行了整合，并参考了 Genevestigator（https：www.genevestigator.ethz.ch gv index.jsp）中的表达数据。

2 结果与讨论

2.1 拟南芥中 PHD-finger 蛋白的确定

为获得拟南芥中所有 PHD-finger 蛋白的信息，我们先从 SMART 中找出了拟南芥中所有预测的 PHD-finger 结构域的序列，共 217 个，其中可能包括了错误的和冗余的部分（Bienz，2006）。为得到更为可靠的结果，我们利用 Pfam 中结构域识别工具对这 217 个结构域序列做了进一步分析，得到 88 个候选 PHD-finger 结构域的序列。我们又根据 Aasland 等（1995）提出的 PHD-finger 结构域的共有序列特征对候选序列做了进一步的筛选，发现其中 84 个结构域序列均为典型的 PHD-finger 结构域，这 84 个 PHD-finger 结构域分属 70 个基因，其中包括了 MMD1、MS1、VIM1 等 24 个已知基因。这 70 个基因编码的蛋白的长度从 234 - 2242 个氨基酸不等，有 14 个基因编码的蛋白拥有 2 个 PHD-finger 结构域（表 1）。

2.2 拟南芥 PHD-finger 蛋白系统进化分析

为了了解拟南芥中 PHD-finger 蛋白间的进化关系，我们用 ClustalX1.83 对 70 个 PHD-finger 蛋白的全序列做了比对，然后根据序列比对结果构建了 NJ（Neighbor-Joining）系统进化树，自举值（bootstrap）设为 1000 次重复，用于检测系统进化树各部分的置信度。结果显示拟南芥 PHD-finger 家族进化树中存在 17 个有较高置信度的进化分枝（自举值 > 900），根据这一结果我们从 PHD-finger 家族成员中分离出 17 个主要的亚家族（命名从 A 到 Q）（图 1），但由于树的其它部分节点处自举值较低，仍有 15 个成员不能划分到任何亚家族。

表 1 拟南芥 PHD-finger 家族成员

Table 1 Members of Arabidopsis PHD-finger family

位点 Locus	别名 Other names	结构域数[a] Domain number	5′端（bp）[b] 5′End	3′端（bp）[c] 3′End	蛋白长度（Aa） Protein length
At1g05380		1	1，576，862	1，582，613	1138
At1g05830	ATX2	1	1，754，243	1，761，816	1056
At1g14510		1	4，964，423	4，961，881	252
At1g14770		1	5，088，841	5，086，621	429
At1g32810		1	11，881，790	11，885，823	1068
At1g33420		1	12，123，556	12，120，866	697
At1g43770		1	16，550，881	16，552，991	431
At1g50620		1	18，752，205	18，755，006	629
At1g54390		1	20，310，691	20，308，234	328
At1g57800	VIM5	1	21，415，948	21，412，288	660
At1g57820	VIM1	1	21，421，611	21，417，835	645
At1g63490		1	23，555，695	23，548，321	1116
At1g66040	VIM4	1	24，590，344	24，587，483	622
At1g66050	VIM2	1	24，593，150	24，596，410	623
At1g66170	MMD1	1	24，645，157	24，642，348	704

续表1

位点 Locus	别名 Other names	结构域数[a] Domain number	5端 (bp)[b] 5′End	3端 (bp)[c] 3′End	蛋白长度 (Aa) Protein length
At1g68030		1	25, 505, 207	25, 504, 041	314
At1g77250		2	29, 028, 368	29, 025, 159	522
At1g77800		1	29, 258, 290	29, 265, 511	1423
At1g79350		1	29, 858, 307	29, 849, 548	1299
At2g01810		1	347, 536	350, 178	697
At2g02470		1	652, 579	654, 914	256
At2g19260		1	8, 363, 223	8, 367, 251	631
At2g25170	CHR6, PKL	1	10, 721, 485	10, 731, 065	1384
At2g27980		1	11, 926, 818	11, 920, 706	1037
At2g31650	ATX1	1	13, 469, 258	13, 462, 349	1062
At2g36720		2	15, 400, 150	15, 406, 497	1007
At2g37520		2	15, 756, 866	15, 751, 847	854
At3g01460	MBD9	2	173, 323	182, 461	2176
At3g02890		1	641, 050	645, 310	963
At3g05670		1	1, 653, 270	1, 657, 165	883
At3g07780	OBERON1	1	2, 484, 475	2, 487, 030	566
At3g08020		2	2, 561, 808	2, 557, 443	779
At3g11200		1	3, 510, 633	3, 508, 048	246
At3g14740		1	4, 953, 430	4, 952, 011	343
At3g14980		2	5, 044, 818	5, 039, 938	1189
At3g19510	HAT3.1	1	6, 766, 570	6, 762, 770	723
At3g20280		1	7, 071, 199	7, 074, 647	743
At3g24010		1	8, 678, 227	8, 675, 965	234
At3g42790		1	14, 890, 769	14, 888, 946	250
At3g52100		2	19, 330, 321	19, 333, 751	696
At3g53680		1	19, 908, 390	19, 903, 841	839
At3g61740		2	22, 867, 947	22, 861, 812	1018
At4g10600		1	6, 548, 860	6, 547, 782	237
At4g12620	ATORC1B	1	7, 462, 250	7, 459, 643	813
At4g14700	ATORC1A	1	8, 422, 226	8, 424, 755	809
At4g22140		1	11, 730, 521	11, 727, 738	234
At4g27910		2	13, 894, 700	13, 900, 480	1027
At4g29940	PRHA	1	14, 653, 385	14, 648, 145	796
At4g39100	SHL1	1	18, 220, 276	18, 217, 933	228
At5g05610		1	1, 679, 105	1, 676, 933	241
At5g09790	ATXR5	1	3, 041, 045	3, 038, 979	379
At5g12400		1	4, 013, 814	4, 020, 910	1595
At5g15540		1	5, 057, 414	5, 047, 676	1846
At5g16680		1	5, 473, 509	5, 467, 537	1290
At5g20510		1	6, 942, 910	6, 939, 815	260
At5g22260	MS1	1	7, 370, 298	7, 367, 638	672
At5g22760		2	7, 571, 638	7, 577, 869	1566
At5g24330	ATXR6	1	8, 297, 071	8, 295, 150	349
At5g26210		1	9, 160, 386	9, 158, 323	255
At5g35210		2	13, 499, 468	13, 491, 426	1576
At5g36670		1	14, 418, 721	14, 423, 657	1193
At5g36740		1	14, 478, 020	14, 482, 957	1179
At5g39550	VIM3	1	15, 857, 864	15, 854, 477	617
At5g44800	CHR4	1	18, 110, 679	18, 100, 646	2242
At5g48160	OBERON2	1	19, 547, 542	19, 545, 059	574
At5g53430	ATX5	2	21, 694, 372	21, 700, 720	1043
At5g57380	VIN3	1	23, 266, 002	23, 263, 621	620
At5g58610		2	23, 708, 279	23, 703, 997	1065
At5g60410	SIZ1	1	24, 312, 116	24, 318, 373	885
At5g63900		2	25, 585, 924	25, 587, 597	557

[a]蛋白质中 PHD-finger 结构域的数目，由 SMART 和 Pfam 鉴定得到。[b]基因的5′端在染色体上的位置。[c]基因的3′端在染色体上的位置。

[a] The number of PHD-finger domain in a protein is determined by both SMART and Pfam searches. [b] The position of 5′ end of gene on chromosomes. [c] The position of 3′ end of gene on chromosomes.

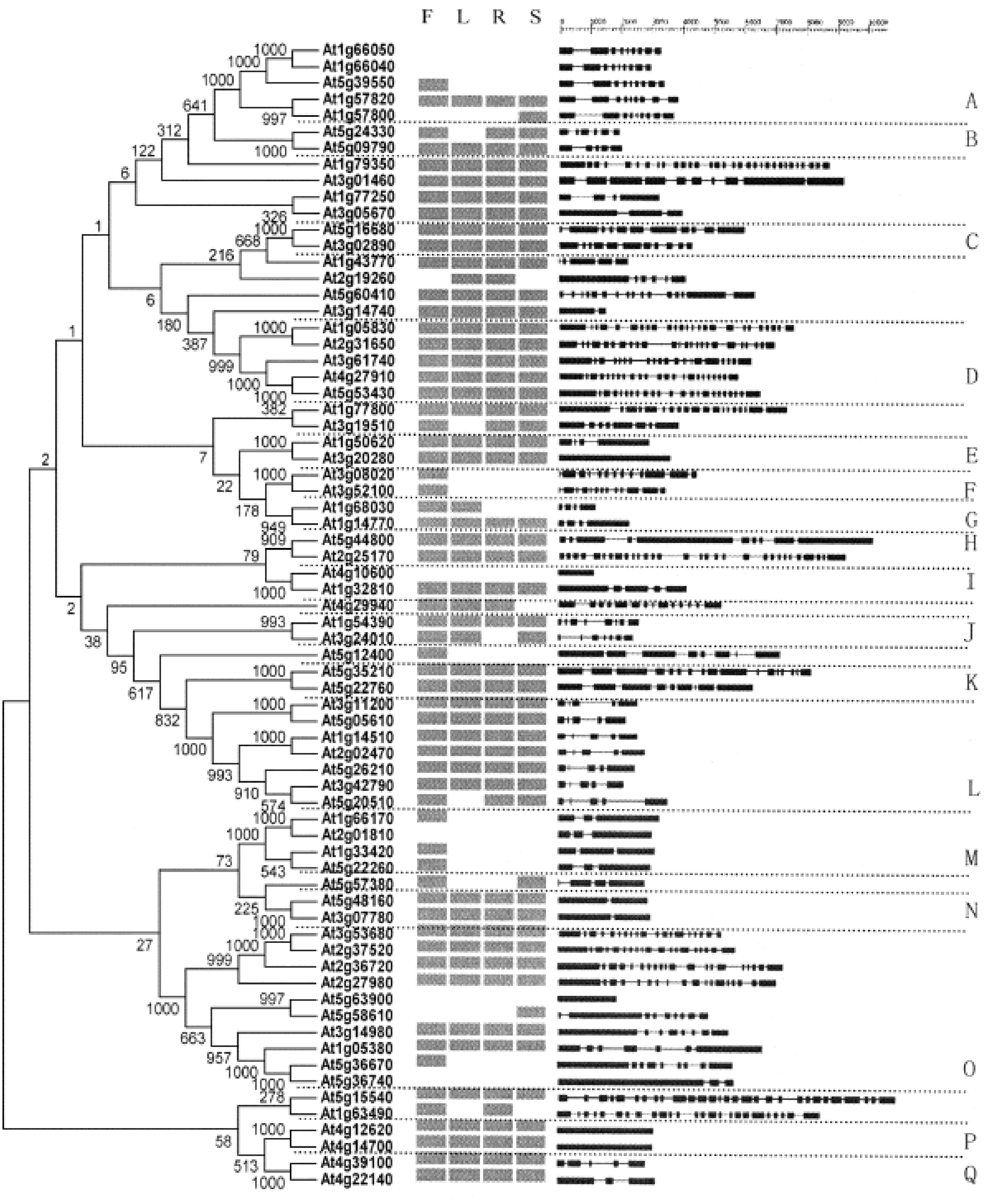

图 1　拟南芥 PHD-finger 蛋白的 NJ 系统进化树及基因表达类型和基因结构

左图为用 ClustalX1.83 **构建的** PHD-finger **蛋白的** NJ **系统进化树，概括了拟南芥** PHD-finger **家族** 70 **个蛋白间的进化关系。蛋白按基因号命名。进化树中用虚线分割出** 17 **个自举值大于** 900 **的亚家族（命名从** A-Q)。**分支旁的数字代表基于** 1000 **次重复的自举值。中间为基因表达类型图，其上方的** F **表示花组织，**R **表示根组织，**L **表示叶组织，**S **表示果荚。每个字母下方的灰色方块表示对应的基因在此组织中表达。右图为基因结构图，黑色方块表示外显子，黑线表示内含子。**

Fig. 1　Structures and expression patterns of the PHD-finger gene and NJ phylogenetic tree of the PHD-finger proteins in *Arabidopsis*

The figure on the left shows the neighbor-joining phylogenetic tree constructed using ClustalX1.83, summarizing the evolutionary relationship among the 70 members of the PHD-finger proteins in *Arabidopsis*. The proteins are named according to their gene loci (see Table 1). 17 subfamilies (named with letters A to Q) with high bootstrap values (＞900) are separated by dashed lines from the phylogenetic tree. The numbers beside the branches represent bootstrap values based on 1000 replications. The middle figure shows the expression patterns of the PHD-finger gene. The letter F above the column of expression data refers to flower, R refers to root, L refers to leaf, and S refers to silique. The positive signals are indicated by gray boxes below the letters. The figure on the right shows the gene structures. The black boxes indicate exons. The black lines between the black boxes indicate introns.

此外我们还通过 NCBI 中 BLASTp 搜索了拟南芥中 PHD-finger 蛋白在水稻中的直系同源蛋白情况。结果正如我们所预料，拟南芥中大多数 PHD-finger 亚家族的成员在水稻中都有同源蛋白，这表明这些蛋白可能在单子叶植物和双子叶植物分离之前已经出现了。而亚家族 F、G、和 I 在水稻中却没有一致的序列，表明这三个亚家族的蛋白分离时间较晚，可能只在双子叶植物中出现。

2.3　拟南芥中 PHD-finger 结构域序列分析

为进一步了解拟南芥中 PHD-finger 结构域间的相似性，我们又比对了 84 个 PHD-finger 结构域的序列（图 2）。这 84 个序列长度从 40～60 个氨基酸不等，图中被阴影突出的区域是与 Zn 离子结合的 7 个半胱氨酸和 1 个组氨酸残基，由于其涉及到结构域的稳定性，因此非常保守，此外其两侧的残基也有一定的保守性，这可能与识别不同甲基化状态的组蛋白有关（Org 等，2008）。而两个 Loop 区是不同序列之间差异最大的部分，不仅长度不同，而且氨基酸残基组成也不同，这些区域与蛋白的功能有密切的关系。根据序列的比对结果，我们发现拟南芥中 PHD-finger 结构域的共有序列为 $C\text{-}X_{(1\text{-}2)}\text{-}C\text{-}X_{(8\text{-}19)}\text{-}C\text{-}X_{(2\text{-}4)}\text{-}C\text{-}X_{(4\text{-}6)}\text{-}H\text{-}X_2\text{-}C\text{-}X_{(11\text{-}26)}\text{-}C\text{-}X_{(2\text{-}3)}\text{-}C$ 与前人的研究结果（Capili 等，2001）基本一致。对于变化较大的两个 Loop 区序列，由于不同作者分析的对象不同，得到的长度范围也有所不同，这也暗示两个 Loop 区不仅在同一物种的不同 PHD-finger 蛋白中差异较大，而且在不同物种的 PHD-finger 蛋白进化中也是变化较大的。

根据 PHD-finger 结构域的多序列比对结果我们构建了结构域的 NJ 进化树（图 3），并从进化树中分离出 23 组进化关系相近的 PHD-finger 结构域，大部分 PHD-finger 结构域的进化关系与蛋白全序列之间的进化关系是一致的，这表明我们对 PHD-finger 蛋白亚家族的划分是较为可靠的，但也存在一些有趣的差异，如：D 亚家族中包含了 5 个 Trithorax 同源蛋白，虽然 5 个蛋白的全序列亲缘关系相近但按其 PHD-finger 结构域间的相似度却能分成两组；不属于同一亚家族的 At3g14740、At3g19510 和 At4g29940 编码的三个蛋白，虽然全序列亲缘关系较远，但 PHD-finger 结构域序列却非常相似等。这些结果指出了同一亚家族成员之间的差异性以及不同亚家族成员之间的相似性，为进一步研究 PHD-finger 蛋白结构与功能之间的关系提供了帮助。

2.4　PHD-finger 家族基因的定位和重复分析

利用 TAIR 网站的染色体绘图工具（Chromosome Map Tool）我们将预测的 PHD-finger 家族 70 个基因定位到 5 条染色体上（图 4）。从图中可以看出，这一家族有 30% 的基因分布在第 5 条染色体上，27% 的基因分布在第 1 条染色体上，21% 的基因分布在第 3 条染色体上，但在第 2 条和第 4 条染色体上这一家族基因分布较少，特别是第 4 条染色体短臂上完全没有这一家族的基因分布。

在植物中，很多基因家族通过大范围的基因组倍增或串联重复方式来扩充家族成员数量（Cannon 等，2004）。为了解拟南芥 PHD-finger 家族成员的进化机制，我们研究了涉及这一家族成员的基因组倍增和串联重复情况。首先我们利用 “Paralogous in *Arabidopsis*” 对 70 个预测的 PHD-finger 蛋白基因进行了分析，发现共有 25 个 PHD-finger 蛋白的基因位于已知的倍增片段上（Blanc 等，2003），其中 20 个基因的倍增基因仍为 PHD-finger 家族基因，并且成对的倍增基因位于 NJ 进化树中相同的进化分枝，但另外 5 个基因的倍增基因中则没有检测到有典型的 PHD-finger 结构域编码序列，其中 *At*4*g*14920、*At*4*g*04260、*At*2*g*18880 都是因为碱基突变导致 PHD-finger 结构域中保守氨基酸发生变化，而 *At*4*g*31900 和 *At*4*g*30200 则是整个 PHD-finger 结构域的缺失，可能是在倍增或重组过程中发生丢失。在分析串联重复基因时我们发现拟南芥 PHD-finger 家族中涉及串联重复的基因比较少，只有第一条染色体长臂上存在两对串联重复基因 *At*1*g*57800（VIM5）和 *At*1*g*57820（VIM1）以及 *At*1*g*66040（VIM4）和 *At*1*g*66050（VIM2），有意思的是这两对基因都属于 A 亚家族。以上结果表明了在拟南芥 PHD-finger 家族中成员的扩增主要依赖于基因组的片段倍增。

2.5　基因结构和蛋白结构域分析

利用 TIGR（http:　www.tigr.org tdb e2k1 ath1）提供的数据，我们对拟南芥 PHD-finger 家族成员的基因结构做了分析。结果表明在这一家族中最长的基因有 10 033 bp，最短的只有 1 078 bp。有 5 个基因没有内含子，而 65 个有内含子的基因，

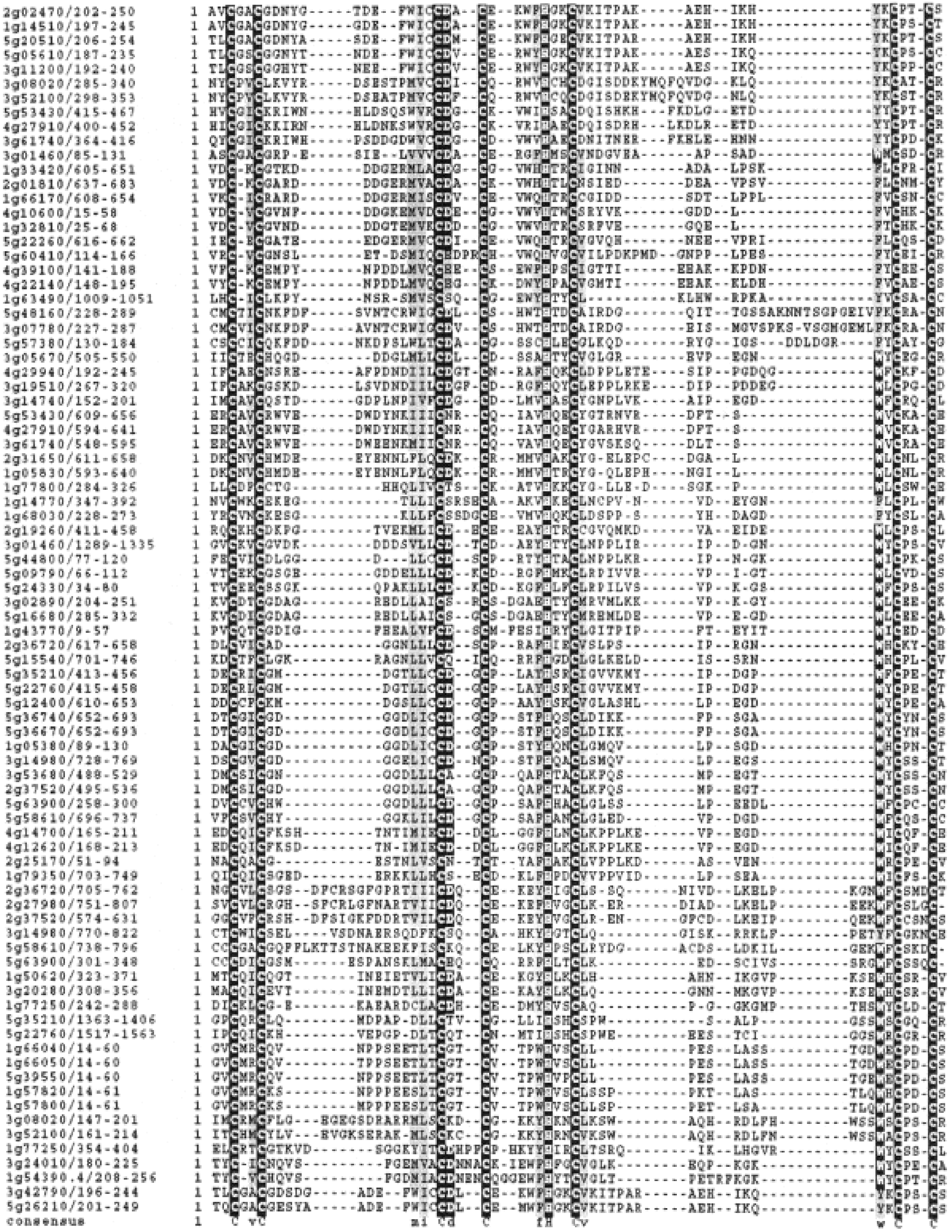

图 2　PHD-finger 结构域序列的多序列比对

对文中确定的 84 个 PHD-finger 结构域序列做比对，结构域序列命名参考所在基因号，结构域名后标注了 PHD-finger 在蛋白中的具体位置。比对中一致的氨基酸残基用黑色阴影标记，相似的氨基酸残基用灰色阴影标记。图底部显示了通过比对得到的保守氨基酸残基。

Fig. 2　Multiple sequence alignment of the PHD-finger domains

Multiple sequence alignment of 84 PHD-finger domains identified in this study. The domains are named according to their gene loci. The positions of the PHD-finger domains in the proteins are indicated by the number at the end of the domain names. The shading of the alignment presents identical residues in black and similar residues in gray. The consensus residues are shown at bottom.

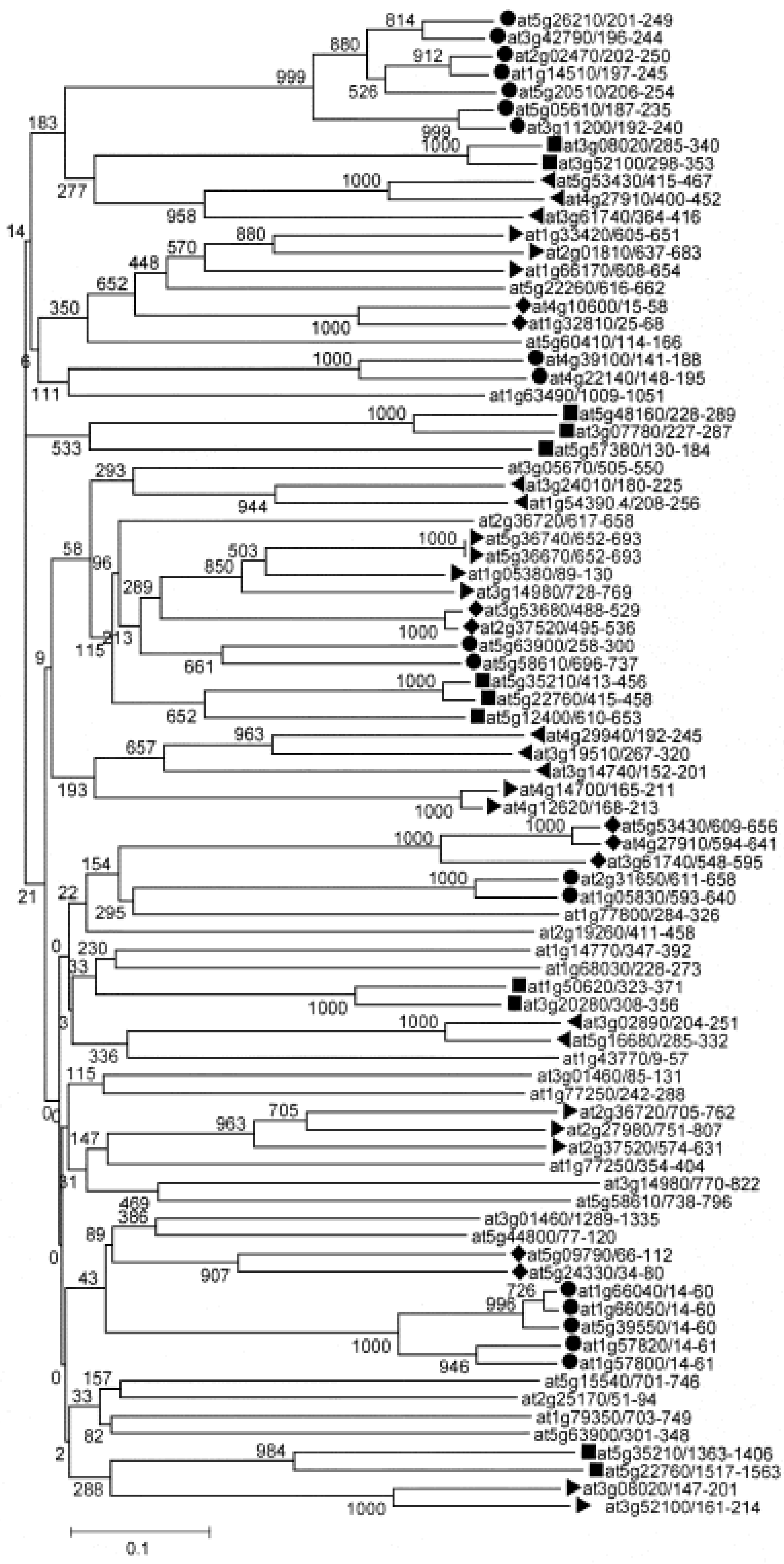

图 3 拟南芥 PHD-finger 结构域的 NJ 系统进化树

进化树的构建基于 PHD-finger **结构域序列的比对结果，概括了拟南芥中** PHD-finger **结构域的进化关系，结构域按基因号命名，并标明结构域在蛋白中的位置，分支旁的数字表示基于** 1000 **次重复的自举值。在一个进化分枝中用相同符号标记的结构域具有较近的进化关系。**

Fig. 3 NJ phylogenetic tree of the PHD-finger domains in *Arabidopsis*

The phylogenetic tree, based on multiple sequence alignment of PHD-finger domains, summarizes the evolutionary relationships among the PHD-finger domains in *Arabidopsis*. The domains are named according to their gene loci and the positions of the PHD-finger domains in the proteins are indicated at the end of the domain names. The numbers beside the branches represent bootstrap values based on 1000 replications. The domains given the same symbols in a clade have close evolutionary relationships.

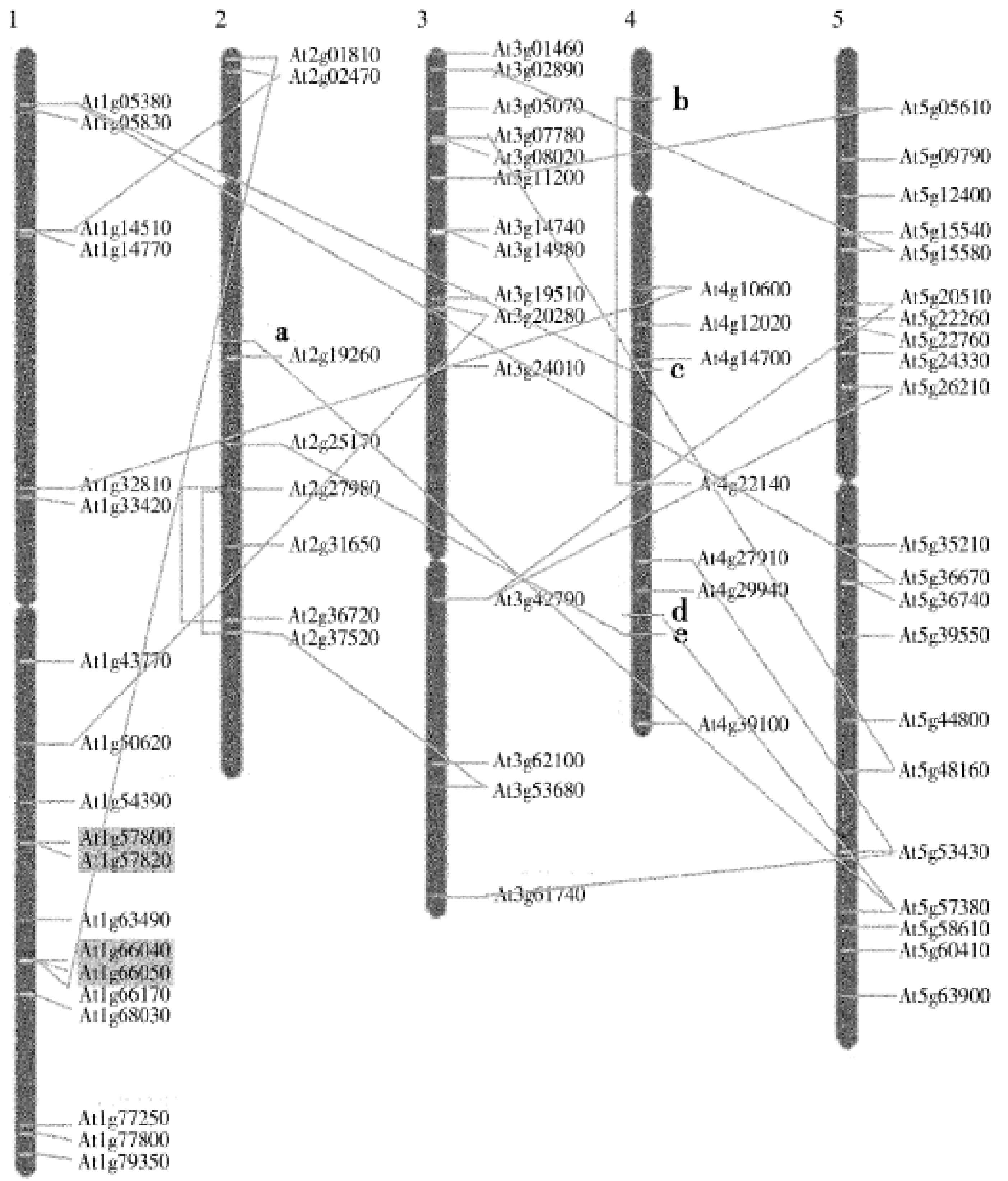

图 4　拟南芥 PHD-finger 基因的分布和倍增情况

图上方标注了染色体号，图中 a、b、c、d、e **分别表示不属于** PHD-finger **家族的** *At2g*18880、*At4g*04260、*At4g*14920、*At4g*30200、*At4g*31900。**倍增基因用灰线相连。串联重复基因被灰色背景标记。**

Fig. 4　Chromosomal distribution and duplication events for *Arabidopsis* PHD-finger genes

The chromosome number is indicated at the top of each chromosome. a: *At2g*18880; b: *At4g*04260; c: *At4g*14920; d: *At4g*30200; e: *At4g*31900. The five genes indicated with letters a to e are not PHD-finger genes. Gray lines connect corresponding duplicated genes. The genes involved tandem duplication events are marked with the gray background.

内含子数目从 1～29 不等（图 1），并且结构相似的基因位于 NJ 进化树中相同的进化分枝，这一结果证明了我们对亚家族的划分是较为可靠的。在对拟南芥 PHD-finger 结构域中内含子分析时发现有 47 个结构域中没有内含子，约占总数的 56%，只有 37 个结构域有内含子，内含子数从 1～3 个不等，大多分布在两个 Loop 中，且 0 相位的内含子居多。

蛋白质的结构域与功能有密切的关系，因此对拟南芥 PHD-finger 蛋白结构域研究有重要意义。我们首先用 MEME 对 70 个 PHD-finger 蛋白分别做了分析，结果表明同一亚家族成员之间含有非常

相似的保守结构基序且很多 PHD-finger 亚家族成员具有 1 个或多个基序完全一致的旁系同源蛋白，如：At1g66050 和 At1g66040；At3g07780 和 At5g48160；At2g36720、At2g37520 和 At3g53680 等，这些旁系同源蛋白之间序列相似并可能存在功能冗余，这就使得人们可能难以通过研究单基因突变体的方式来分析这些基因的功能。而大多数不在同一亚家族的 PHD-finger 蛋白之间则差异较大，不仅 PHD-finger 以外的基序序列差异较大，不同基序的位置也各不相同。

同时我们又用 SMART 和 Pfam 结构域分析工具鉴定并注释了 PHD-finger 蛋白中除 PHD-finger 结构域之外的已知结构域（图 5），在 A 亚家族中我们发现所有成员都有 2 个 RING 结构域和 1 个 SRA 结构域，SRA 结构域可能与 DNA 中的甲基化胞嘧啶相结合（Unoki 等，2004）。而 B 亚家族和 D 亚家族中都含有 SET 结构域，SET 结构域可能参与组蛋白的甲基化过程（Wang 等，2001），此外 D 亚家族成员还都具有 PWWP 和 PostSET 结构域。H 亚家族中的成员都有 CHROMO、DEADc 和 HELICc 结构域，而 K、P、Q 三个亚家族成员分别都具有 DDT、AAA 和 BAH 结构域。在其它亚家族和 15 个单独的 PHD-finger 蛋白中还鉴定到一些有趣的结构域，如：JmjC、HOX、FN3 等。根据 SMART 和 Pfam 的鉴定结果我们发现大多数 PHD-finger 蛋白的功能结构域都与染色质结构调控以及基因的转录调控有关，这为我们进一步研究 PHD-finger 家族的生物学功能提供了帮助。

2.6 PHD-finger 家族基因表达分析

通常基因的表达谱与基因的功能有关，我们主要利用了 MPSS 和 ESTs 数据库分析了 PHD-finger 家族基因的表达情况，对于两个数据库中不一致的，我们将二者信息进行整合，并参考了 Genevestigator 数据库的信息。从表达谱中可以看出 PHD-finger 家族大多数基因在各个组织中均有表达（图 1），表明这些基因可能普遍参与拟南芥各个组织的生长发育，但也有一些基因的表达具有组织特异性，如包括 MMD1 和 MS1 两个已知基因在内的 M 亚家族成员主要在花中表达，MMD1 的突变体表现为小孢子母细胞减数分裂异

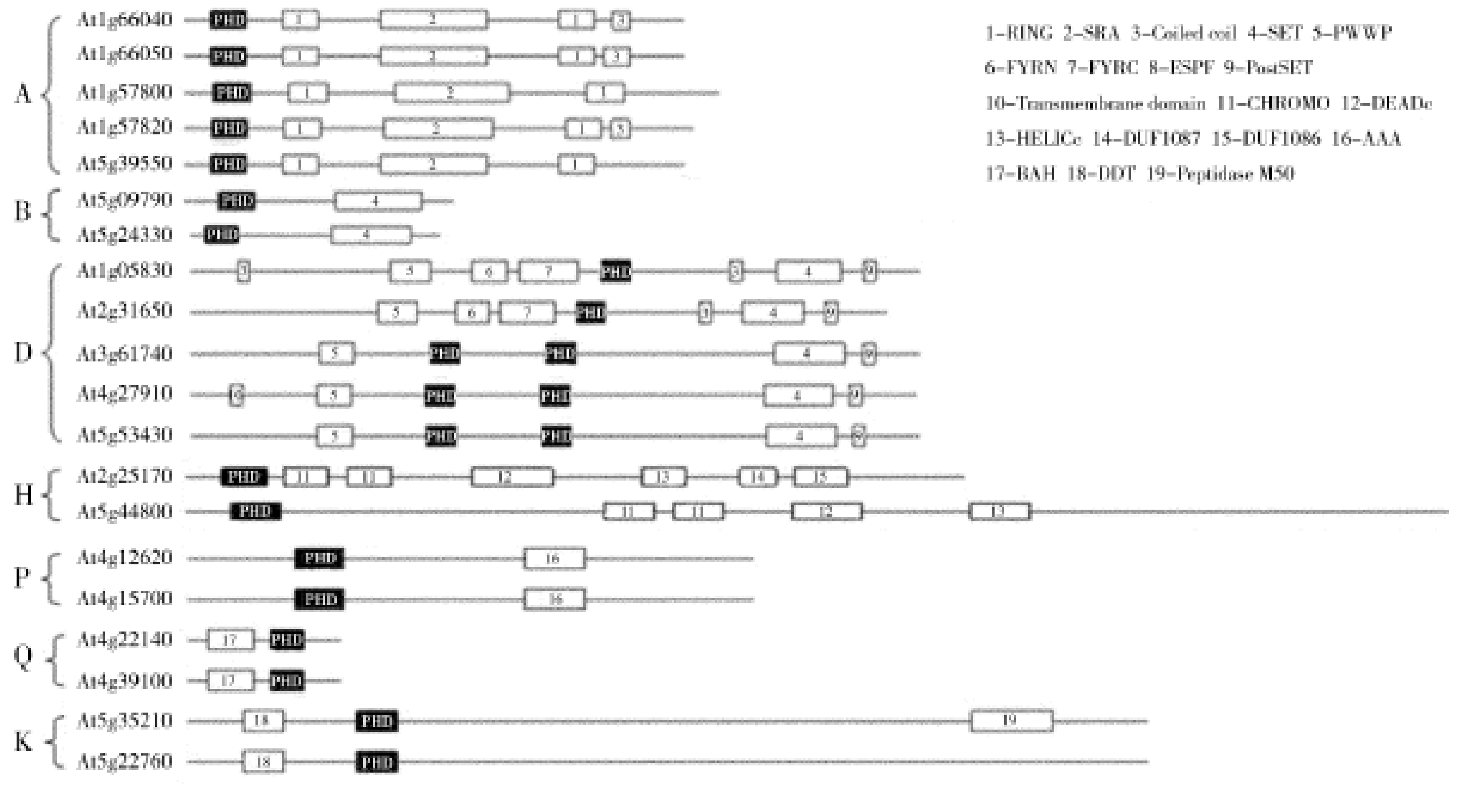

图 5 部分亚家族蛋白的结构域

图中列出了 A、B、D、H、K、P、Q 亚家族蛋白的 PHD-finger 结构域及其它已知结构域，结构域由 SMART 和 Pfam 鉴定获得。黑色方块表示 PHD-finger 结构域，用数字标记的白方块表示其它结构域，其具体类型显示在右上方。

Fig. 5 Domains of protein in some subfamily

The protein domains in A，B，D，H，K，P and Q Subfamily determined by both SMART and Pfam searches are shown. The black boxes indicate the PHD-finger domain. The white boxes with numbers 1 to 19 indicate 19 domains except PHD-finger domain. The names of the 19 domains are shown in top right.

常并导致雄性不育（Yang等，2003），而MS1的突变体表现为绒毡层和小孢子外壁异常（Yang等，2007）。此外F亚家族以及*At5g*12400等基因也在花中特异性表达。

大多数亚家族成员的表达谱具有一致性，但也有的亚家族成员的表达谱差异较大，如A亚家族中*At1g*57820（VIM1）表达最为广泛，而其它VIM基因则表达各异，这一家族成员主要参与结合DNA中甲基化胞嘧啶和组蛋白，并调控染色质的状态，很明显VIM1在其中起了主要作用，研究表明VIM1的突变体会导致DNA高度甲基化并使处于分裂间期的染色质凝缩，而其它几个基因的突变体则没有明显表型变化（Woo等，2007）。O亚家族成员的表达谱也不尽相同，但基因结构和蛋白结构域相近的*At3g*53680、*At2g*37520、*At2g*36720和*At2g*27980 4个基因表达却较为一致。

尽管拟南芥PHD-finger家族大部分成员的功能未知，但系统进化和表达分析的结果为今后深入的结构和功能研究提供了参考。

3 结论

对拟南芥中PHD-finger家族的70个基因做了较全面的研究，包括分别对结构域和蛋白全序列的比对及系统进化树的构建，基因的倍增和染色体定位分析，基因结构及表达情况和蛋白结构域的分析等。根据拟南芥PHD-finger家族蛋白进化树的结果，我们分离出17个进化关系较近的亚家族，亚家族划分的可靠性得到了其它分析结果的支持。在PHD-finger结构域序列的分析中我们揭示了相同亚家族成员和不同亚家族成员之间PHD-finger结构域的相似和差异，为深入研究PHD-finger蛋白功能提供了参考。对基因组倍增和串联重复的研究使我们认识到基因组片段倍增事件在PHD-finger家族成员数的扩增中起了主要的作用。而基因表达及蛋白结构域的分析则揭示了PHD-finger家族蛋白可能在拟南芥不同组织的生长发育中必不可少，它们大多参与染色体状态的调控和基因的转录调控等。总之以上结果为我们进一步研究PHD-finger家族成员的结构和功能的关系提供了参考。

〔参 考 文 献〕

Aasland R, Gibson TJ, Stewart AF, 1995. The PHD finger: implications for chromatin mediated transcriptional regulation [J]. *Trends in Biochemical Sciences*, 20 (2): 56—59

Bienz M, 2006. The PHD finger, a nuclear protein-interaction domain [J]. *Trends in Biochemical Sciences*, 31 (1): 35—40

Blanc G, Hokamp K, Wolfe KH, 2003. A recent polyploidy superimposed on older large-scale duplications in the *Arabidopsis* genome [J]. *Genome Research*, 13 (2): 137—144

Cannon SB, Mitra A, Baumgarten A *et al.*, 2004. Young ND, May G. The roles of segmental and tandem gene duplication in the evolution of large gene families in *Arabidopsis thaliana* [J]. *BMC Plant Biology*, 4: 10

Capili AD, Schultz DC, RauscherⅢ FJ *et al.*, 2001. Solution structure of the PHD domain from the KAP-1 corepressor: structural determinants for PHD, RING and LIM zinc-binding domains [J]. *The EMBO Journal*, 20: 165—177

Eberharter A, Vetter I, Ferreira R *et al.*, 2004. ACF1 improves the effectiveness of nucleosome mobilization by ISWI through PHD-histone contacts [J]. *The EMBO Journal*, 23 (20): 4029—4039

Feng Y (冯英), Liu QP (刘庆坡), Xue QZ (薛庆中), 2004. Comparative phylogenetic analysis of the rice and *Arabidopsis* PHD-finger proteins [J]. *Acta Genetica Sinica* (遗传学报), 31 (11): 1284—1293

Gibbons RJ, Bachoo S, Picketts DJ *et al.*, 1997. Mutations in transcriptional regulator ATRX establish the functional significance of a PHD-like domain [J]. *Nature Genetics*, 17 (2): 146—148

Kaadige MR, Ayer DE, 2006. The polybasic region that follows the plant homeodomain zinc finger 1 of Pf1 is necessary and sufficient for specific phosphoinositide binding [J]. *Journal of Biological Chemistry*, 281 (39): 28831—28836

Kehle J, Beuchle D, Treuheit S *et al.*, 1998. dMi-2, a hunchback-interacting protein that functions in polycomb repression [J]. *Science*, 282 (5395): 1897—1900

Kramps T, Peter O, Brunner E *et al.*, 2002. Wnt wingless signaling requires BCL9 legless-mediated recruitment of pygopus to the nuclear beta-catenin-TCF complex [J]. *Cell*, 109 (1): 47—60

Lan F, Collins RE, Cegli RD *et al.*, 2007. Recognition of unmethylated histone H3 lysine 4 links BHC80 to LSD1-mediated gene repression [J]. *Nature*, 448 (7154): 718—723

Martin DG, Baetz K, Shi XB *et al.*, 2006. The Yng1p Plant Homeodomain Finger Is a Methyl-Histone Binding Module That Recognizes Lysine 4-Methylated Histone H3 [J]. *Molecular and Cellular Biology*, 26 (21): 7871—7879

Nakamura Y, Umehara T, Hamana H *et al.*, 2007. Crystal structure analysis of the PHD domain of the transcription co-activator pygopus [J]. *Journal of Molecular Biology*, 370 (1): 80—92

Ogryzko VV, Schiltz RL, Russanova V *et al.*, 1996. The transcriptional coactivators p300 and CBP are histone acetyltransferases [J]. *Cell*,

87 (5): 953—959

Org T, Chignola F, Hetényi C *et al.*, 2008. The autoimmune regulator PHD finger binds to non-methylated histone H3K4 to activate gene expression [J]. *The EMBO Journal*, 9 (4): 370—376

Papoulas O, Beek SJ, Moseley SL *et al.*, 1998. The Drosophila trithorax group proteins BRM, ASH1 and ASH2 are subunits of distinct protein complexes [J]. *Development*, 125 (20): 3955—3966

Saleh A, Alvarez-Venegas R, Yilmaz M *et al.*, 2008. The highly similar *Arabidopsis* homologs of trithorax ATX1 and ATX2 encode proteins with divergent biochemical functions [J]. *Plant Cell*, 20 (3): 568—579

Schultz DC, Friedman JR, Rauscher Ⅲ FJ, 2001. Targeting histone deacetylase complexes via KRAB-zinc finger proteins: the PHD and bromodomains of KAP-1 form a cooperative unit that recruits a novel isoform of the Mi-2a subunit of NuRD [J]. *Genes* & *Development*, 15 (4): 428—443

Shi XB, Kachirskaia L, Walter KL *et al.*, 2007. Proteome-wide analysis in saccharomyces cerevisiae identifies several PHD fingers as novel direct and selective binding modules of histone H3 methylated at either lysine 4 or lysine 36 [J]. *Journal of Biological Chemistry*, 282 (4): 2450—2455

Talatsuji H, 1998. Zinc-finger transcription factors in plants [J]. *Cellular and Molecular Life Sciences*, 54 (6): 582—596

Townsley FM, Thompson B, Bienz M, 2004. Pygopus residues required for its binding to Legless are critical for transcription and development [J]. *Journal of Biological Chemistry*, 279 (7): 5177—5183

Unoki M, Nishidate T, Nakamura Y, 2004. ICBP90, an E2F-1 target, recruits HDAC1 and binds to methyl-CpG through its SRA domain [J]. *Oncogene*, 23 (46): 7601—7610

Wang H, Cao R, Xia L *et al.*, 2001. Purification and functional characterization of a histone H3-lysine 4-specific methyltransferase [J]. *Molecular Cell*, 8 (6): 1207—1217

Woo HR, Pontes O, Pikaard CS *et al.*, 2007. VIM1, a methylcytosine-binding protein required for centromeric heterochromatinization [J]. *Genes* & *Development*, 21 (3): 267—277

Wysocka J, Swigut T, Xiao H *et al.*, 2006. A PHD finger of NURF couples histone H3 lysine 4 trimethylation with chromatin remodelling [J]. *Nature*, 442 (7098): 86—90

Yang C, Vizcay-Barrena G, Conner K *et al.*, 2007. MALE STERILITY1 is required for tapetal development and pollen wall Biosynthesis [J]. *Plant Cell*, 19 (11): 3530—3548

Yang X, Makaroff CA, Ma H, 2003. The *Arabidopsis* MALE MEIOCYTE DEATH1 gene encodes a PHD-finger protein that is required for male meiosis [J]. *Plant Cell*, 15 (6): 1281—1295